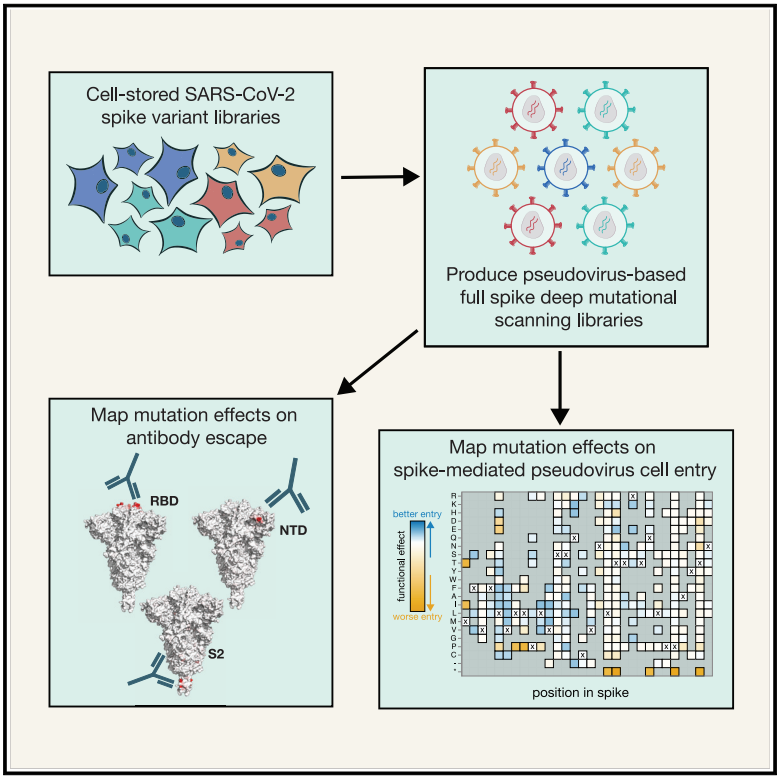
Cell

Resource

A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike

Graphical abstract



Authors

Bernadeta Dadonaite, Katharine H.D. Crawford, Caelan E. Radford, ..., Helen Y. Chu, Richard A. Neher, Jesse D. Bloom

Correspondence

jbloom@fredhutch.org

In brief

A high-throughput deep mutational scanning platform that uses non-replicative pseudotyped lentiviruses is able to directly quantify how SARS-CoV-2 spike mutations affect antibody neutralization and spike-mediated infection.

Highlights

- Method for genotype-phenotype-linked lentivirus pseudotyping
- Make full SARS-CoV-2 spike deep mutational scanning libraries
- Use libraries to map antibody-escape mutations across spike
- Measure effects of mutations in spike on virus cell entry

Dadonaite et al., 2023, Cell *186*, 1263–1278
March 16, 2023 © 2023 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2023.02.001

CellPress

Cell

CellPress
OPEN ACCESS

Resource

A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike

Bernadeta Dadonaite,[1,15] Katharine H.D. Crawford,[1,2,15] Caelan E. Radford,[1,3,15] Ariana G. Farrell,[1] Timothy C. Yu,[1,3] William W. Hannon,[1,3] Panpan Zhou,[4,5,6] Raiees Andrabi,[4,5,6] Dennis R. Burton,[4,5,6,7] Lihong Liu,[8] David D. Ho,[8,9,10] Helen Y. Chu,[11] Richard A. Neher,[12,13] and Jesse D. Bloom[1,14,16,*]

[1]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[2]Department of Genome Sciences & Medical Scientist Training Program, University of Washington, Seattle, WA 98109, USA
[3]Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA 98109, USA
[4]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[5]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA
[6]Consortium for HIV/AIDS Vaccine Development (CHAVD), The Scripps Research Institute, La Jolla, CA 92037, USA
[7]Ragon Institute of Massachusetts General Hospital, MIT & Harvard, Cambridge, MA 02139, USA
[8]Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA
[9]Department of Microbiology and Immunology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA
[10]Division of Infectious Diseases, Department of Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA
[11]University of Washington, Department of Medicine, Division of Allergy and Infectious Diseases, Seattle, WA, USA
[12]Biozentrum, University of Basel, Basel, Switzerland
[13]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[14]Howard Hughes Medical Institute, Seattle, WA 98195, USA
[15]These authors contributed equally
[16]Lead contact
*Correspondence: jbloom@fredhutch.org
https://doi.org/10.1016/j.cell.2023.02.001

SUMMARY

A major challenge in understanding SARS-CoV-2 evolution is interpreting the antigenic and functional effects of emerging mutations in the viral spike protein. Here, we describe a deep mutational scanning platform based on non-replicative pseudotyped lentiviruses that directly quantifies how large numbers of spike mutations impact antibody neutralization and pseudovirus infection. We apply this platform to produce libraries of the Omicron BA.1 and Delta spikes. These libraries each contain ~7,000 distinct amino acid mutations in the context of up to ~135,000 unique mutation combinations. We use these libraries to map escape mutations from neutralizing antibodies targeting the receptor-binding domain, N-terminal domain, and S2 subunit of spike. Overall, this work establishes a high-throughput and safe approach to measure how $\sim 10^5$ combinations of mutations affect antibody neutralization and spike-mediated infection. Notably, the platform described here can be extended to the entry proteins of many other viruses.

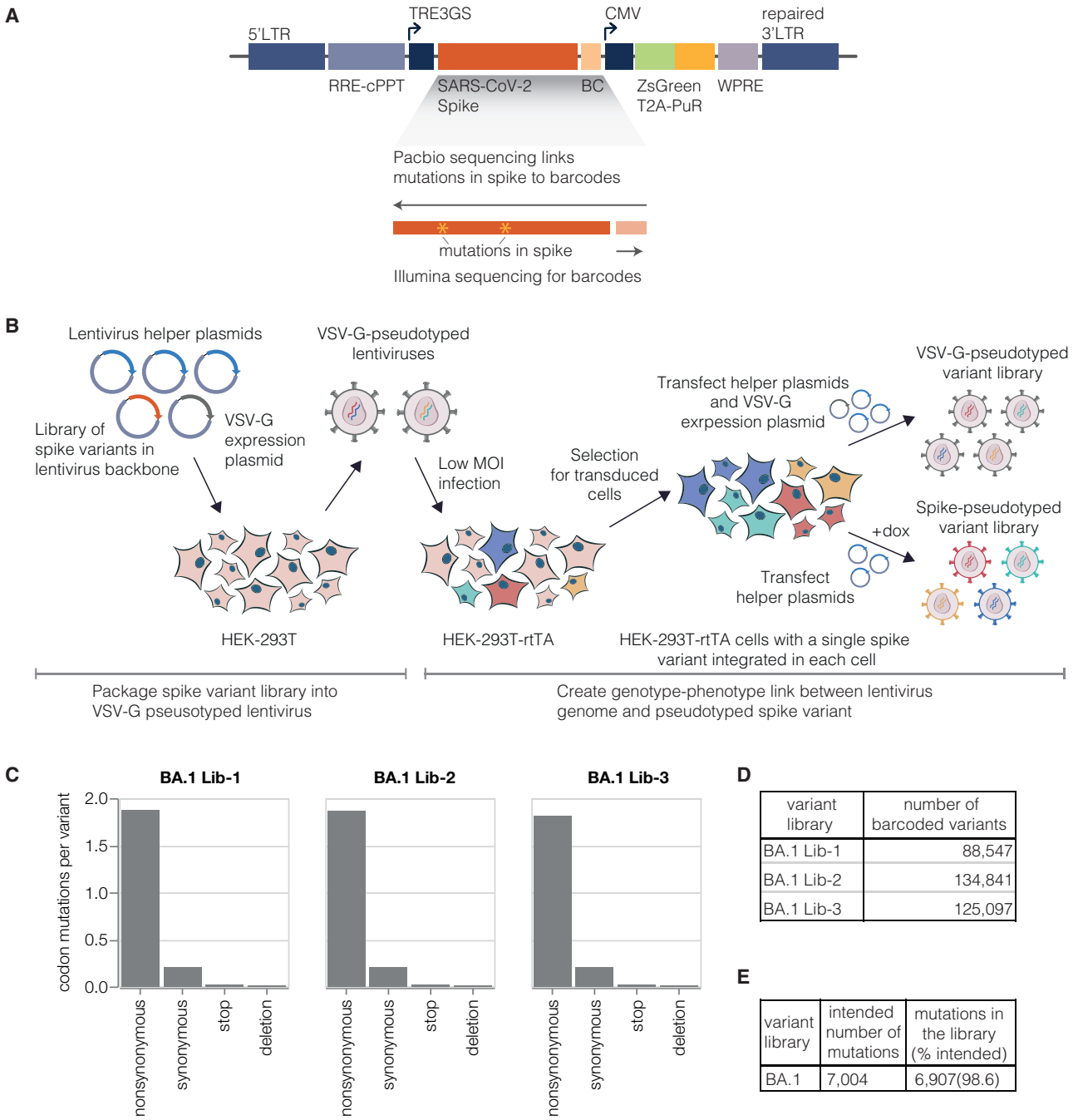
INTRODUCTION

The spike protein is the key target of neutralizing antibodies against SARS-CoV-2. Unfortunately, spike has undergone rapid evolution, which has eroded the potency of serum neutralization and enabled escape from most monoclonal antibodies.[1–4] Deep mutational scanning experiments can prospectively measure the effects of large numbers of mutations even before they emerge in viral variants, and therefore, they have been a valuable tool for rapidly interpreting how newly observed mutations in the spike affect antibody binding and protein folding or function.[1,5,6] The high-throughput nature of deep mutational scanning experiments has also enabled the generation of huge datasets that can inform computational methods for predicting the antigenic properties of possible future viral variants.[1,7]

However, prior deep mutational scanning of the SARS-CoV-2 spike has been limited to either solely focusing on the receptor-binding domain (RBD),[1,8,9] other subdomains,[10,11] or just a small number of mutations across spike.[12] Furthermore, all previous spike deep mutational scanning experiments have been based on cell-surface display using either yeast[8,13] or mammalian cells[10–12] and therefore are limited to measuring antibody binding rather than neutralization, despite the fact that neutralization is thought to be a more relevant correlate of protection.[14,15]

Here, we describe a deep mutational scanning platform that directly measures how mutations affect cellular infection and

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



D

variant library	number of barcoded variants
BA.1 Lib-1	88,547
BA.1 Lib-2	134,841
BA.1 Lib-3	125,097

E

variant library	intended number of mutations	mutations in the library (% intended)
BA.1	7,004	6,907(98.6)

Figure 1. Deep mutational scanning platform for spike

(A) Lentivirus backbone used for deep mutational scanning. The backbone contains functional lentiviral 5′ and 3′ long terminal repeat (LTR) regions. The spike gene is under an inducible tetresponse element 3rd generation (TRE3G) promoter, and there is a 16-nucleotide barcode (BC) downstream of the stop codon. A CMV promoter drives the expression of the reporter ZsGreen gene that is linked to a puromycin resistance gene (PuR) via a T2A linker. The backbone also contains a woodchuck hepatitis virus posttranscriptional regulatory element (WPRE), Rev response element (RRE), and a central polypurine tract (cPPT).

(B) Approach for creating genotype-phenotype linked lentivirus libraries. HEK-293T cells are transfected with spike-carrying lentivirus backbone, VSV-G expression plasmid, and lentiviral helper plasmids to generate VSV-G-pseudotyped lentiviruses. These viruses are used to transduce reverse tetracycline-controlled transactivator (rtTA) expressing HEK-293T cells at low multiplicity of infection (MOI), and successfully transduced cells are selected using puromycin. Selected cells can be transfected with helper plasmids and a VSV-G expression plasmid to produce VSV-G-pseudotyped viruses carrying all genomes present in

(legend continued on next page)

antibody neutralization in the context of the full SARS-CoV-2 spike pseudotyped on non-replicative lentiviral particles. The key innovation behind the platform is a two-step pseudovirus generation protocol that enables the creation of large pseudovirus libraries with a link between the lentiviral genotype and the particular spike protein variant on the pseudovirus surface. We demonstrate that this platform can be used to create large genotype-phenotype linked pseudovirus libraries and map how mutations to spike affect both cellular infection and neutralization by antibodies targeting diverse regions of spike, including the RBD, N-terminal domain (NTD), and S2 subunit.

RESULTS

Producing pseudoviruses with genotype-phenotype link

To characterize thousands of mutations in spike glycoprotein, we first established a lentiviral pseudotyping platform that maintains a genotype-phenotype link between the lentiviral genome and the spike variant on the virion's surface. Lentiviral spike-pseudotyping usually involves transfection of a backbone that carries a reporter gene flanked by the lentiviral long terminal repeats (LTRs), helper plasmids that code for structural and nonstructural genes required for the lentiviral life cycle, and an expression plasmid that codes for the spike variant of interest (Figure S1A).[16–18] When these components are transfected into producer cells, virions are formed that carry lentiviral genomes and display spikes on their surface. However, because genome incorporation into a virion does not depend on the expressed spike, there is no link between the virion's genotype and the phenotype of the spike on its surface (Figure S1A). The absence of a genotype-phenotype link is not problematic when only a single spike variant is used for transfection; however, it precludes deep mutational scanning studies that involve studying thousands of variants in a single pooled experiment.

To create a lentiviral genotype-phenotype link, we first generated a lentivirus backbone with the following key elements (Figure 1A): (1) we restored the ability of the lentivirus to transcribe its full genome after integration by repairing the 3′ LTR deletion present in traditional lentivirus vectors,[19,20] (2) we placed spike in the lentivirus backbone under an inducible promoter, and (3) we added a second constitutive promoter driving both a fluorescent reporter (ZsGreen) and a puromycin resistance gene.

Next, we developed a multi-step protocol that creates a genotype-phenotype link by ensuring that each producer cell only expresses a single variant of spike (Figure 1B). In the first step of this protocol, we transfected cells with the spike-encoding backbone, a vesicular stomatitis virus G protein (VSV-G) expression plasmid, and the necessary helper plasmids. This produces non-genotype-phenotype-linked VSV-G-pseudotyped lentiviruses that we use to infect target cells at low multiplicity of infection so that most infected cells receive no more than one lentiviral genome. Next, we select for cells with integrated lentiviral genomes using puromycin, which yields a population of cells where each cell stores only a single spike variant. The spike is under an inducible promoter, which is only activated by the addition of doxycycline. To produce virions, we induce spike expression with doxycycline and transfect the helper plasmids necessary to produce lentiviruses. We validated that this approach can be used to generate genotype-phenotype-linked spike-pseudotyped viruses with titers $>10^5$ transduction units per mL (Figure S1B). We can further increase viral titers by ~5- to 10-fold by infecting cells in the presence of amphotericin B, as has been reported previously (Figure S1C).[21–23]

Design of mutations in SARS-CoV-2 spike deep mutational scanning library

Rather than create deep mutational scanning libraries containing all possible amino acid mutants of spike, we chose to introduce only mutations that seem likely to arise during natural evolution and yield a functional spike protein. We had two rationales for designing our libraries in this way: (1) it reduces the total number of mutations that need to be included in the library, and (2) it increases the probability that variants with multiple mutations will remain functional by reducing the fraction of mutations that are highly deleterious.

Specifically, we included only mutations that have been observed in spike sequences deposited on the GISAID database,[24] reasoning that these mutations would represent mostly functional spike proteins. We introduced mutations at a higher frequency when they have emerged in spike independently many times according to the pre-built SARS-CoV-2 phylogenies from UShER.[25] Finally, we included every possible amino acid change at sites in spike that are evolving under positive selection.[26] We also included deletions at sites where such mutations are observed frequently in natural SARS-CoV-2 evolution. In total, our library design targeted 7,004 mutations in the BA.1 spike and 6,852 mutations in the Delta spike.

To introduce these mutations in the spike gene, we used a PCR-based mutagenesis method with a primer pool containing the desired mutations.[27] Importantly, this method introduces multiple mutations in each spike variant: we targeted ~2 to 3 codon mutations per variant, ensuring the effects of most mutations are measured in multiple genetic backgrounds. The mutated spike genes were then barcoded with 16 random nucleotides placed downstream of the spike-coding sequence (Figure 1A) and cloned into the lentivirus backbone. As described below, after integration of the libraries into cells, these barcodes can be linked to the full set of mutations in each spike variant to facilitate downstream sequencing.[28,29]

the deep mutational library or selected cells can be induced with doxycycline (dox) to express spike and transfected with only the helper plasmids to generate spike-pseudotyped lentiviruses that have a genotype-phenotype link.
(C) Average number of mutations per barcoded spike in BA.1 libraries.
(D) Total number of barcoded variants in each BA.1 library.
(E) The coverage of intended mutations across all BA.1 libraries.
See also Figures S1, S2, S3, S6, and S7.

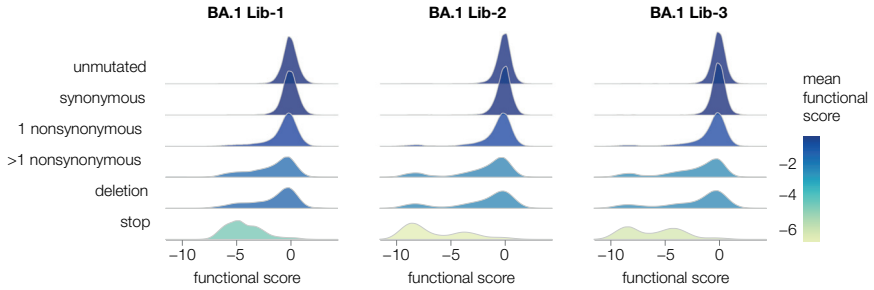


Figure 2. Some mutations tend to impair spike-mediated pseudovirus infection

For each barcoded spike variant, we compute a functional score that reflects how well that spike mediates pseudovirus infection relative to the unmutated spike: negative scores indicate impaired infection, and positive scores indicate improved infection. The plots show the distribution of functional scores across all variants in each of the three BA.1 libraries for different categories of variants, with each distribution colored by the mean functional score for that variant type.

See also Figure S2.

Production of pseudotyped BA.1 and Delta spike deep mutational scanning libraries

We used the genotype-phenotype linked pseudovirus production strategy in Figure 1B to make BA.1 and Delta deep mutational scanning libraries. We created three independent BA.1 libraries each containing ~100,000 barcoded variants and two independent Delta libraries each containing ~50,000 barcoded variants (Figures 1E and S2A; a "barcoded variant" is a spike with a unique nucleotide barcode and some random mutation set; different barcoded variants usually but not always contain different mutations). After integrating the libraries into cells at low multiplicity of infection, we generated VSV-G-pseudotyped lentivirus from these cells by co-transfecting a plasmid expressing VSV-G alongside the other lentiviral helper plasmids (Figure 1B, top right). The use of a VSV-G-pseudotyped virus ensures that we generate infectious lentiviral virions from all integrated backbones regardless of whether they encode a functional spike mutant. We then infected this VSV-G-pseudotyped lentivirus into a new round of cells and performed long-read PacBio sequencing to link the barcodes to the full set of spike mutations for each variant. We performed the PacBio barcode-mutation linking after integration into cells because recombination of the pseudodiploid lentiviral genome during integration[30,31] means the barcode-mutation pairings may be different in the integrated cells to those in the original lentiviral backbone plasmids.[32] Importantly, linking barcodes to spike variants allows us to use short-read Illumina sequencing of the barcode to obtain the full spike genotype in all subsequent experiments.

Overall, the sequencing revealed that we had successfully introduced ~99% of the targeted mutations in the BA.1 and Delta spike libraries (Figures 1E and S2B), with at least one amino acid mutation at each site in spike for both the BA.1 and Delta libraries. The barcoded variants in the BA.1 libraries had on average ~2 codon mutations per spike, whereas the variants in the Delta libraries had ~3 codon mutations per spike (Figures 1C and S2C). The number of mutations per variant is roughly Poisson distributed, so some variants had zero or one mutation, whereas others had many more (Figure S3).

We then generated the actual spike-pseudotyped deep mutational scanning libraries from the variants stored at a single copy in the cells (Figure 1B, lower right). We calculated a functional score for each variant based on its relative frequency in the spike- versus VSV-G-pseudotyped libraries. Positive functional scores indicate spike variants mediate pseudovirus infection better than the parental spike, whereas negative functional scores indicate worse pseudovirus infection. As expected, spike variants with premature stop codons had highly negative functional scores, whereas unmutated and synonymously mutated spike variants had functional scores close to zero (Figures 2, S2D, and S2E). Some variants with nonsynonymous mutations had functional scores close to zero, whereas others had more negative scores, reflecting the fact that some but not all nonsynonymous mutations are deleterious (Figures 2 and S2D; recall that our library design protocol preferentially introduced nonsynonymous mutations expected to yield functional spikes). Variants with multiple nonsynonymous mutations tended to have lower functional scores than variants with just one nonsynonymous mutation (Figures 2 and S2D), reflecting the cost of accumulating multiple often mildly deleterious mutations. Variants that only contained mutations designed into the libraries tended to have higher functional scores than variants that also carried other unintended nonsynonymous mutations (Figure S2E), suggesting that our library design strategy indeed enriched for functionally tolerated mutations.

Use of an absolute standard to measure viral neutralization by deep sequencing

Traditional neutralization assays measure the infectivity of a single virus variant at multiple antibody concentrations. Deep sequencing can measure the relative infectivities of many viral variants in pooled infections in the presence of an antibody. However, to convert the relative infectivities measured by deep sequencing into actual neutralization values, it is necessary to have an absolute standard that does not vary in its infectivity as a function of antibody concentration (Figure 3A). To enable such measurements in our experiments, we added a barcoded VSV-G-pseudotyped virus into our libraries. This virus is produced from a separate cell line that has integrated lentiviral backbones carrying a barcoded mCherry gene (see STAR Methods). Importantly, this VSV-G-pseudotyped virus is not neutralized by any of the spike-binding antibodies (Figures 3A and S4A), and so the counts for VSV-G barcodes provide an absolute neutralization standard. To calculate the non-neutralized fraction for each viral variant, we simply compute the change in its barcode frequencies relative to the VSV-G standard (Figure 3B).

To validate this approach, we added the VSV-G absolute standard at ~1% of our BA.1 library titers and incubated the

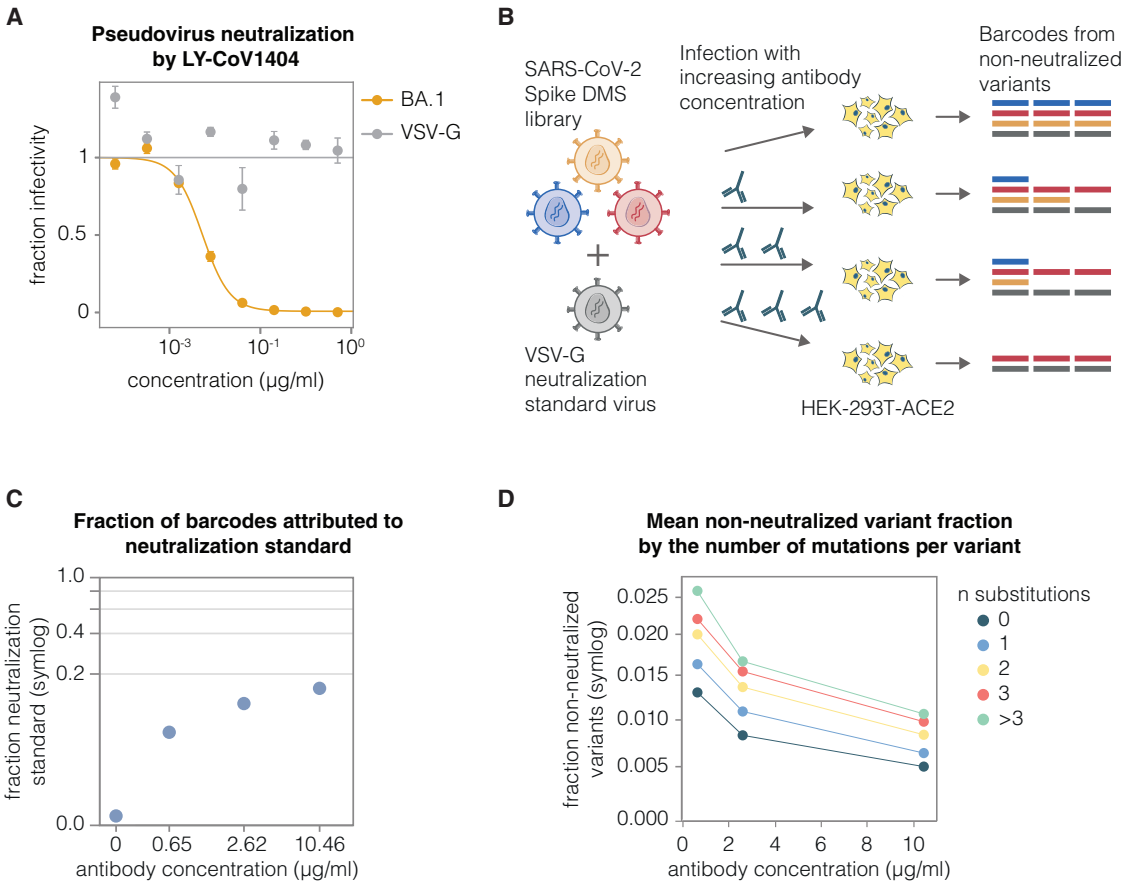


Figure 3. A VSV-G standard enables the measurement of absolute neutralization by deep sequencing

(A) Neutralization assay demonstrating that BA.1-spike-pseudotyped lentivirus is neutralized by antibody LY-CoV1404, but the VSV-G-pseudotyped neutralization standard is not. Error bars indicate standard error between two technical replicates.

(B) Use of the VSV-G standard to measure absolute neutralization. Deep mutational scanning libraries are mixed with the VSV-G neutralization standard. The virus mixture is incubated with a no-antibody control or increasing antibody concentrations and infected into ACE2-expressing 293T cells. After ~12 h, viral genomes are recovered, barcodes are sequenced, and absolute neutralization of each variant is computed by comparing its barcode counts to those from the VSV-G standard.

(C) Fraction of barcodes derived from the VSV-G neutralization standard in infections with increasing LY-CoV1404 concentrations.

(D) BA.1 deep mutational scanning library non-neutralized fractions averaged across variants with different numbers of amino acid mutations at differentLY-CoV1404 concentrations. Note (C) and (D) use a symlog scale.

See also Figures S4, S6, and S7.

virus library with increasing concentrations of the LY-CoV1404 antibody, as schematized in Figure 3B. We then infected the library into ACE2-expressing target cells overnight, recovered viral genomes, and quantified the abundance of each viral barcode using deep sequencing. As expected, the fraction of VSV-G standard reads increased with antibody concentration because fewer spike variants could still infect in the presence of antibodies (Figure 3C). We then calculated the non-neutralized fraction for each viral variant in our libraries after selection at different concentrations of the antibody. As expected, increasing antibody concentrations led to decreased non-neutralized fraction averaged over variants (Figure 3D). Notably, variants with a greater number of substitutions had higher non-neutralized fractions, as expected if some substitutions escape the antibody.

Mapping antibody escape using a full spike deep mutational scanning system

To demonstrate that pseudovirus-based deep mutational scanning can map escape from neutralizing antibodies targeting any region of spike, we chose a set of BA.1-neutralizing antibodies that bind distinct regions of spike: RBD-binding LY-CoV1404, NTD-binding 5–7, and S2-binding CC67.105.[33–35] Note that LY-CoV1404, also known as bebtelovimab, is one of the few clinically approved antibodies that retains potency against BA.1, BA.2, and other major Omicron lineages.[3,34]

We first mapped escape from LY-CoV1404, applying the approach in Figure 3B to our three independent BA.1 libraries, and performing a technical replicate for one library. We used a biophysical model to decompose the measurements for spike variants in our libraries (some of which are multiply mutated)

into escape scores for individual mutations.[36] These mutation-escape scores correlated well among technical and biological replicates (Figure 4A). As expected, the key LY-CoV1404 escape sites were in the antibody's previously described epitope in the RBD,[34] which spans sites 439–452 and 498–501 (Figures 4B–4D and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/LyCoV-1404_escape_plot.html). However, our deep mutational scanning emphasizes that only some mutations at these sites escape LY-CoV1404 neutralization. For instance, many amino acid mutations at site 446 strongly escape LY-CoV1404, but mutating this site from G (the identity in Wuhan-Hu-1) to S (the identity in BA.1 and BA.2.75) does not have a large effect. This observation emphasizes the somewhat serendipitous nature of the preserved potency of LY-CoV1404. However, this antibody may soon be escaped because sub-variants of BA.5 and BA.2.75 with mutations in the key escape site of K444 are increasingly being detected.[1,37]

To validate the LY-CoV1404 deep mutational scanning, we cloned a set of mutations in the BA.1 spike with a range of effects in the deep mutational scanning data and performed standard pseudovirus neutralization assays (Figure 4E). All the tested mutations exhibited neutralization phenotypes consistent with those measured in the deep mutational scanning. Furthermore, the neutralization assay IC_{50} values correlated well with those predicted by our biophysical model[36] parameterized by the deep mutational scanning data (Figure 4F).

We also compared the full spike LY-CoV1404 deep mutational scanning measurements to results from our previously described yeast-display system for deep mutational scanning of only the RBD.[8,13] The escape scores between the two experimental approaches correlated well (Figure S5A) and both methods identified the same epitope (Figure S5B).

To show that we can map escape from non-RBD-targeting antibodies, we next mapped the NTD-targeting 5–7 antibody.[33] This antibody targets an epitope outside the defined antigenic supersite in NTD and is one of the few NTD-targeting antibodies isolated pre-Omicron that still retains some potency against Omicron variants.[2,33,38] The deep mutational scanning showed that the key escape sites for 5–7 were in a hydrophobic pocket next to the N4 loop (site 172–178) (Figures 5A–5C and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/NTD_5-7_escape_plot.html), consistent with prior structural characterization of this antibody's epitope.[33] In addition, deletions in 167–171 ß-sheet, as well as mutations at the base of the adjacent loops such as G103 and V126 also escaped antibody 5–7 (Figures 5A and 5B). We validated these deep mutational scanning results by performing individual neutralization assays with pseudoviruses containing L176K, S172N, and G103F mutations (Figure 5D), all of which had the expected effect of completely escaping neutralization (Figure 5E).

We next applied the full spike deep mutational scanning to S2 stem-helix targeting antibodies CC9.104 and CC67.105, which were isolated using both the SARS-CoV-2 and Middle East respiratory syndrome coronavirus (MERS-CoV) spike proteins as baits.[35] Both CC9.104 and CC67.105 broadly neutralize SARS-related coronaviruses, and CC9.104 also retains some potency against MERS-CoV. As expected, our deep mutational scanning showed that escape sites for both antibodies cluster in the S2 stem-helix region (Figures 6A–6E; https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC67.105_escape_plot.html and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC9.104_escape_plot.html). Our data also explain why only CC9.104 neutralizes MERS-CoV. The deep mutational scanning shows that the CC67.105 epitope centers on sites D1146, D1153, and F1156 (Figures 6B and 6D), and consistent with the deep mutational scanning, mutating these sites leads to complete escape in validation neutralization assays (Figure 6G). By contrast, the deep mutational scanning shows that although CC9.104's epitope also includes sites D1153 and F1156, mutations at site D1146 cause only modest or no escape (Figures 6C and 6E), and validation neutralization assays again confirm these deep mutational scanning results (Figure 6G). Notably, sites D1153 and F1156 are conserved between SARS-CoV-2 and MERS-CoV S2 stem-helix regions, but site D1146 is mutated to isoleucine in MERS-CoV (Figure 6F). Based on our deep mutational scanning data we expect changes at site D1146 to have significant effects on escape from the CC67.105 antibody but only modest effects on escape from the CC9.104 antibody. Although our deep mutational scanning libraries do not include isoleucine at site D1146 (which is found in MERS-CoV), we confirmed that other mutations at D1146 lead to complete escape from CC67.105 and do not substantially impact neutralization by CC9.104 (Figure 6G). Note that site D1163 is also mutated to isoleucine in MERS-CoV and both antibodies show some escape at that site, which may explain why CC9.104's potency against MERS-CoV is lower than against SARS-CoV-2.

The above deep mutational scanning of escape from the S2 antibodies emphasizes the difference between SARS-related coronavirus breadth and resistance to escape in SARS-CoV-2. Both CC9.104 and CC67.105 neutralize many diverse SARS-related coronaviruses, but Omicron sub-variants with mutations that lead to almost complete escape from these antibodies have been detected (e.g., D1153Y in BA.2.46 and BA.2.59). Therefore, even pan-sarbecovirus neutralizing antibodies can be escaped by mutational diversity within SARS-CoV-2, which emphasizes the importance of directly mapping escape mutations in SARS-CoV-2 in addition to assessing breadth across other natural SARS-related coronaviruses.

To show that we can perform deep mutational scanning of the spikes from different SARS-CoV-2 strains, we mapped escape from the REGN10933 antibody using Delta spike deep mutational scanning libraries (Figures S6A and S6B). REGN10933 is a class 1 antibody that directly competes with ACE2 binding and was part of the REGN-COV2 therapeutic cocktail used early in the pandemic but has lost potency against Omicron variants.[2,39,40] Escape sites for REGN10933 mapped with our deep mutational scanning system overlapped with the antibody binding footprint and included previously described escape mutations (Figure S6C)[39,40] and correlated well with previously described yeast-based RBD deep mutational scanning data for this antibody (Figures S5C and S5D).[5]

Finally, to show that pseudovirus-based deep mutational scanning can be used to map escape from polyclonal sera, we mapped escape from two Delta breakthrough sera using our

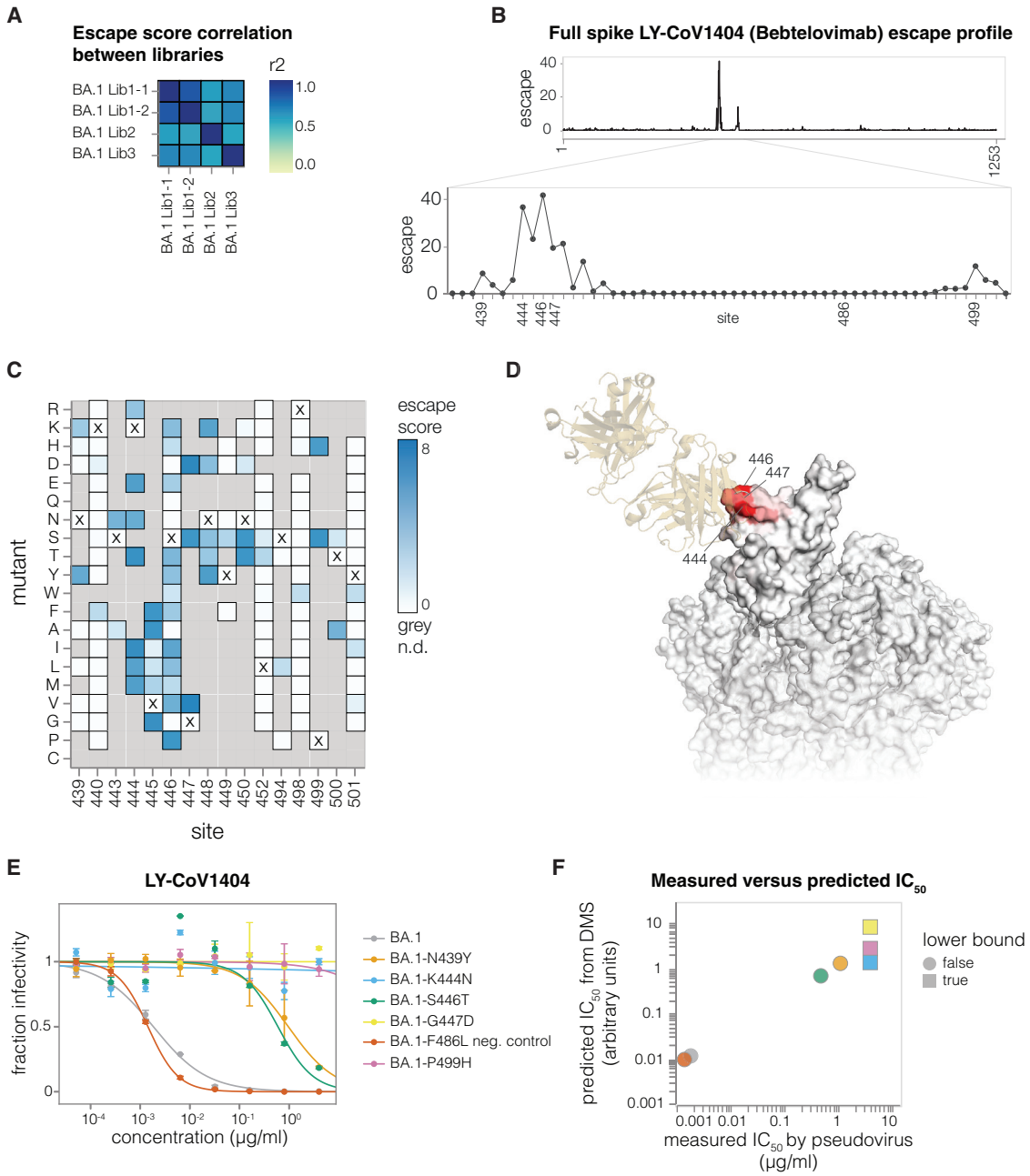


Figure 4. Antibody LY-CoV1404 escape mapping

(A) Correlation of mutation-escape scores between technical replicates (BA1 Lib-1.1 and BA1 Lib-1.2) and biological replicates (BA1 Lib-1, BA1 Lib-2, and BA1 Lib-3).

(B) Total escape scores at each site in the BA.1 spike, and a zoomed-in plot showing the key escape sites. Sites of mutations chosen for validation experiments are labeled on the x-axis.

(C) Heatmap of mutation-escape scores at key sites. Residues marked with X are the wild-type amino acids in BA.1. Amino acids not present in our libraries are shown in gray. An interactive heatmap for the entirety of spike is at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/LyCoV-1404_escape_plot.html.

(D) Surface representation of spike colored by the sum of escape scores at that site. LY-CoV1404 antibody is in yellow. Only the antibody-bound protomer is colored. PDB: 7MMO and PDB: 6XM4 were aligned to make this structure.

(E) Validation pseudovirus neutralization assays of the indicated BA.1 spike mutants against the LY-CoV1404 antibody. Error bars indicate standard error between two technical replicates.

(F) Correlation between predicted IC_{50} values from deep mutational scanning (DMS) data versus the IC_{50} values measured in the validation assays in (E). The points are colored as in (E). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.

See also Figures S4 and S5.

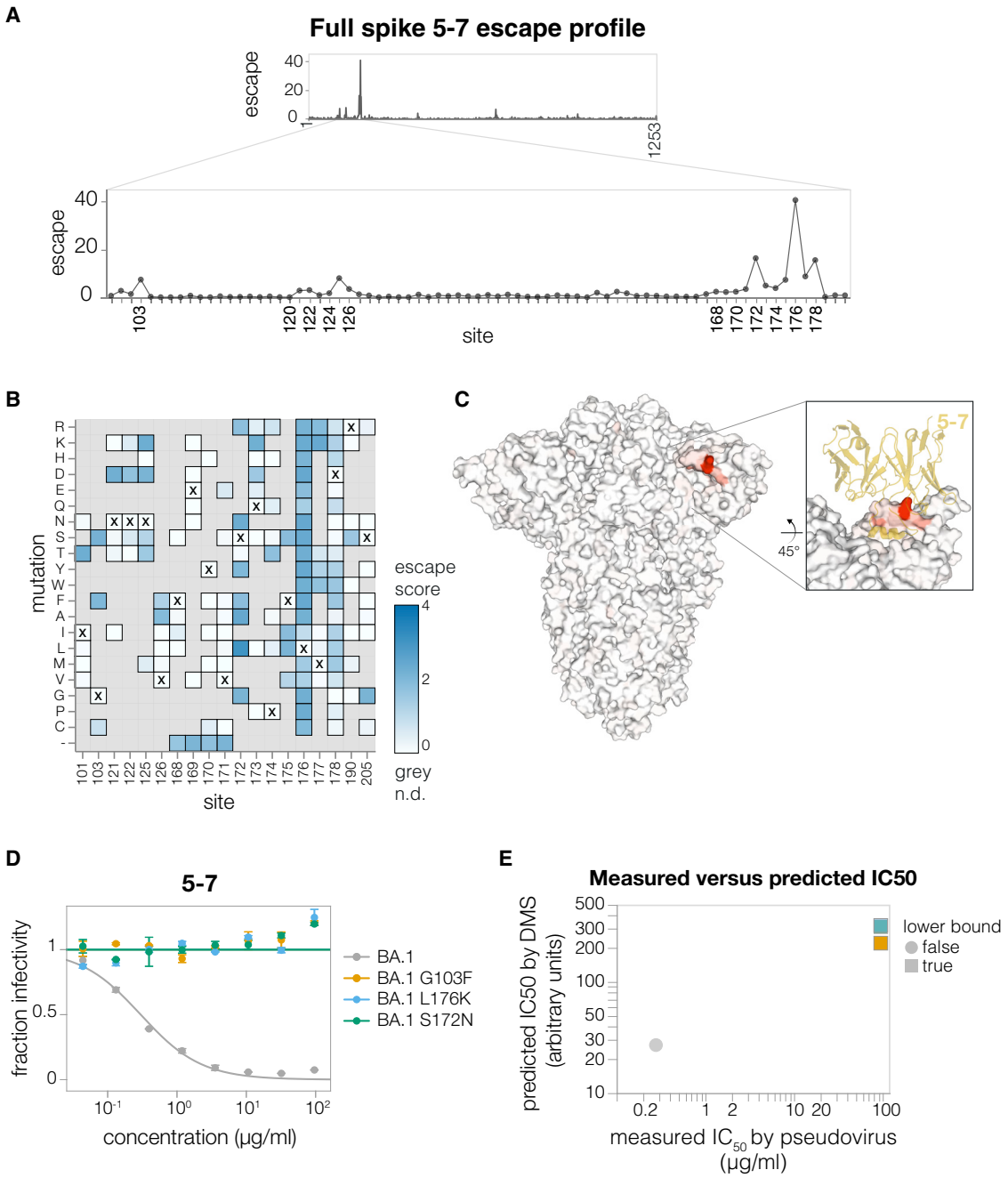


Figure 5. Antibody 5–7 escape mapping

(A) Total escape scores for each site in the BA.1 spike and a zoomed-in plot showing the key escape sites.

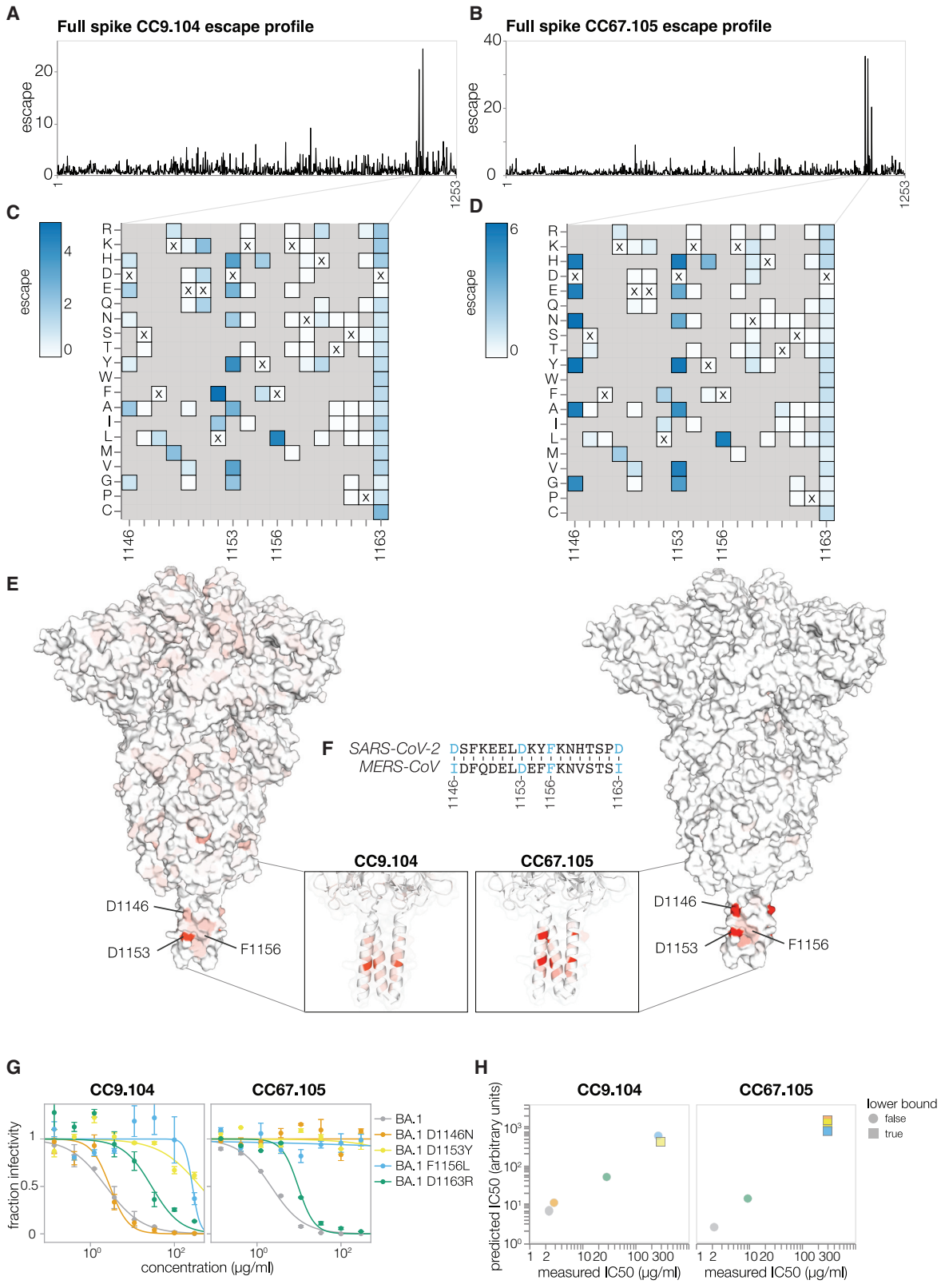
(B) Heatmap of mutation-escape scores at key sites. Residues marked with X are the wild-type amino acids in BA.1. Amino acids not present in our libraries are shown in gray. An interactive version of this plot for the entirety of spike is at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/NTD_5-7_escape_plot.html.

(C) Surface representation of spike colored by the sum of escape scores at that site. Antibody 5–7 is shown in yellow in the inset. PDB: 7RW2.

(D) Validation pseudovirus neutralization assays of the indicated BA.1 spike mutants against antibody 5–7. Error bars indicate standard error between two technical replicates.

(E) Correlation between predicted IC_{50} values from DMS data versus the IC_{50} values measured in the validation assays in (D). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence.

See also Figure S4.



(legend on next page)

Delta libraries (Figure S7). Major escape sites for both sera resided in RBD class I and II epitopes (sites 417, 484, and 486) (Figure S7) as has been shown previously[7] and correlated well with changes in IC50 measured in conventional neutralization assays.

Functional effects of mutations on spike-mediated pseudovirus infection

Our deep mutational scanning also enables measurement of how mutations affect spike-mediated viral infection in the absence of antibodies. We can make these measurements by computing a functional score for each variant from its relative frequency in infectious spike-pseudotyped lentiviruses generated from our single-copy cell integrated cells versus VSV-G-pseudotyped lentivirus generated from the same cells (Figure 1B). Spike variants with negative functional scores are worse at mediating cellular infection than the parental unmutated spike, whereas variants with positive functional scores are better at mediating infection (Figure 2). To deconvolve the functional scores for the variants (which often contain multiple mutations) into the effects of individual mutations on spike-mediated entry, we used global epistasis models.[41,42]

As expected stop codon mutations to the BA.1 spike were highly deleterious for spike-mediated infection, whereas amino acid mutations showed a wide range of effects ranging from slightly beneficial to roughly neutral to highly deleterious (Figure 7A; recall that our library design excludes many of the most deleterious amino acid mutations). To test whether the mutations measured to have slightly beneficial effects actually improved spike-mediated infection, we chose five mutations that the deep mutational scanning indicated improved infection (Figure 7B) and generated pseudovirus mutants carrying these mutations. The validation experiments confirmed that all the tested mutations indeed slightly improved spike-mediated infection (Figure 7C), validating that our deep mutational scanning can identify mutations that increase spike-mediated pseudovirus infection.

To examine the relationship between the functional effects of spike mutations in the deep mutational scanning and the actual evolution of human SARS-CoV-2, we determined the extent that mutations are enriched or depleted across a phylogenetic tree of all publicly available human SARS-CoV-2 sequences.[25] To do this, we calculated the number of independent observations of each mutation on the tree and compared these observed numbers to the expected numbers under neutrality as estimated from 4-fold synonymous sites,[43] analyzing only mutations expected to have $\geq$20 occurrences (see STAR Methods for details). Our deep mutational scanning measurements of the effects of mutations on spike-mediated infection were reasonably correlated with the enrichment of mutations among actual sequences (Figure 7D), indicating our experiments at least partially reflect the functional selection actually shaping spike evolution. We performed a similar analysis for prior spike deep mutational scanning using yeast display of the RBD,[13] or mammalian cell display of the NTD[10] or a region of S2[11] (Figure 7D). Our pseudovirus-based spike deep mutational scanning measurements were more correlated with the enrichment of mutations during actual evolution than any of these prior cell-surface display deep mutational scanning studies, presumably because our experiments mimic the true biological function of spike better than cell-surface display experiments.

However, none of the mutations with positive deep mutational scanning functional scores that we validated to improve spike-mediated infection in a pseudovirus context are enriched during actual SARS-CoV-2 evolution (Figure 7D). We suggest that this is because there is some divergence between the selection pressure in our pseudovirus-based experiments and true natural selection on spike. For instance, mutations at sites P1140 and P1143, which are located at the beginning of the S2 stem-helix, could potentially destabilize the prefusion trimer leading to more rapid cell entry in a pseudovirus context but negatively affect spike stability in the context of actual human transmission. Nonetheless, our functional measurements still provide the most accurate large-scale measurements to date on the effects of mutations to spike and should be useful for assessing which antibody-escape mutations are well enough tolerated to pose a plausible risk of emerging naturally. We also note that our experiments indicate that there are no further mutations to the BA.1 spike that improve pseudovirus titers to the same extent as the D614G mutation that fixed early in SARS-CoV-2's evolution in humans.[44–46]

DISCUSSION

We have developed a deep mutational scanning system for assessing the antigenic and functional effects of mutations in the SARS-CoV-2 spike. This deep mutational scanning system is the first to measure how mutations to the entirety of spike affect

Figure 6. Antibody CC9.104 and CC67.105 escape mapping

(A and B) Total escape scores for each site in the BA.1 spike for the CC9.104 (A) and CC67.105 (B) antibodies.

(C and D) Escape heatmaps for the S2 stem-helix (sites 1146–1163) for CC9.104 (C) and CC67.105 (D) antibodies. Residues marked with X are the wild-type amino acids in the BA.1 sequence. Amino acids that are not present in our libraries are shown in gray. Interactive heatmaps for the entirety of spike are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC67.105_escape_plot.html and https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/CC9.104_escape_plot.html.

(E) Surface representation of spike colored by the sum of escape scores at that site for CC9.104 (left) and CC67.105 (right) antibodies. Site 1163 is not resolved in the structure. PDB: 6XR8.

(F) Alignment of SARS-CoV-2 and MERS-CoV spikes at sites 1146–1163.

(G) Validation pseudovirus neutralization assay for CC9.104 (left) and CC67.105 (right) antibodies with BA.1 spike carrying the indicated mutations. Error bars indicate standard error between two technical replicates.

(H) Correlation between predicted IC_{50} values from DMS data versus the IC_{50} values measured in the validation assays in (G). Lower bound indicates that the antibody did not neutralize at the highest concentration tested in the validation neutralization assay. Site numbering in all plots is based on the Wuhan-Hu-1 sequence. See also Figure S4.

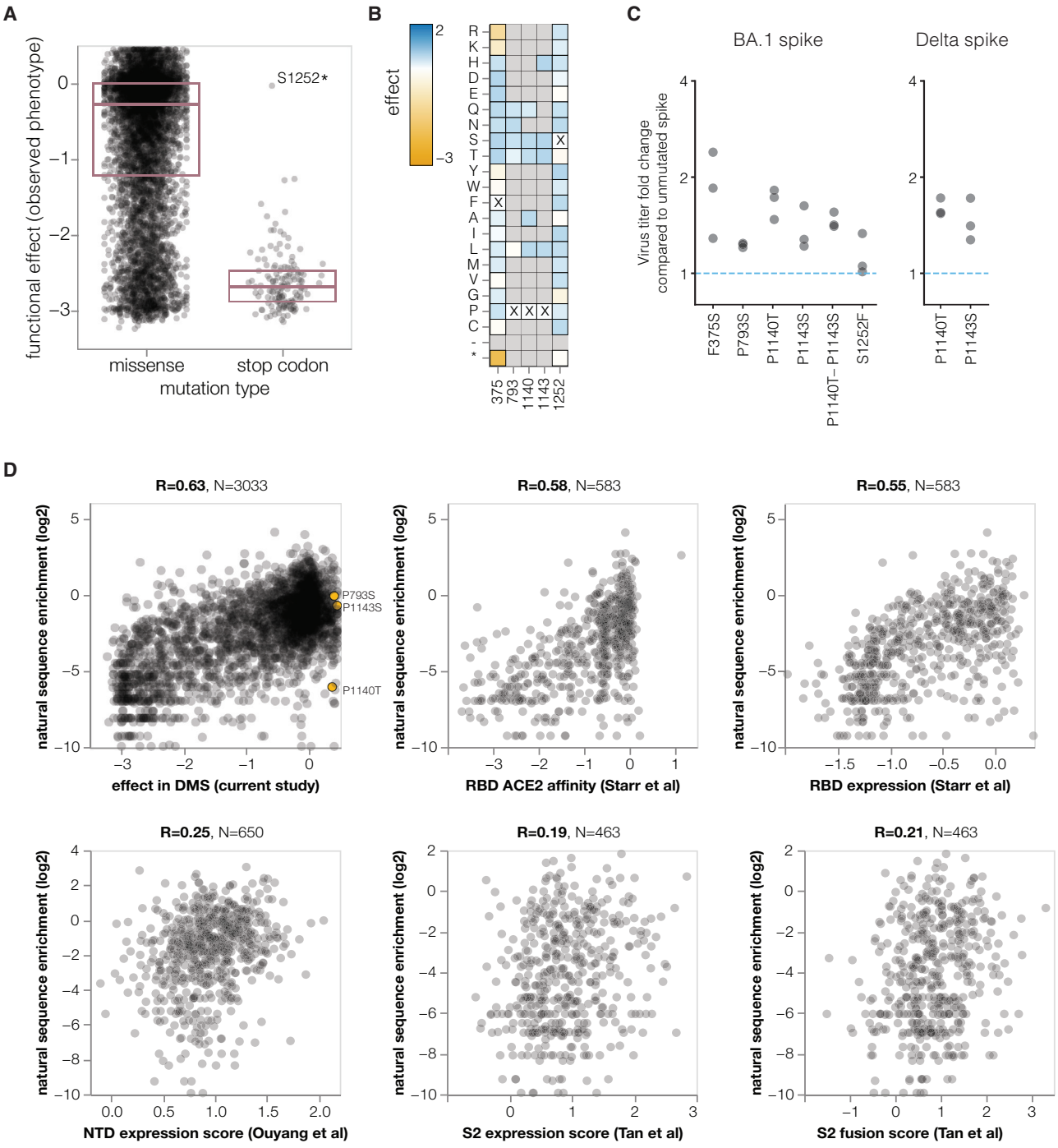


Figure 7. Functional effects of mutations on spike-mediated Pseudovirus infection

(A) Distribution of functional effects of mutations in BA.1 deep mutational scanning libraries. Negative values indicate mutations are deleterious for viral entry. The stop codon mutation with a neutral functional effect of ~0 is at the last codon of the spike used in our experiments.

(B) Heatmap showing functional effects at sites of mutations with beneficial functional effects that were chosen for validation assays in (C). An interactive version of this heatmap for the entire spike is at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/muteffects_observed_heatmap.html.

(C) Fold change in virus entry titer for spike mutants relative to unmutated spike. There are three points for each mutant, reflecting triplicate measurements.

(D) Correlation between enrichment of mutations during actual evolution of human SARS-CoV-2 and functional effects from our lentivirus-based deep mutational scanning or previous RBD expression or ACE2 affinity for yeast-based deep mutational scanning Starr et al.,[13] and S2 Tan et al.[11] or Ouyang et al.[10] expression for

(legend continued on next page)

cellular infection and therefore enables the mapping of escape from antibodies targeting any part of the spike. Furthermore, our system directly quantifies how mutations affect antibody neutralization, and we show that these measurements correlate well with traditional pseudovirus neutralization assays. We expect that the ability to directly measure neutralization as opposed to binding will be especially useful when applied to polyclonal sera since the magnitude of how mutations affect neutralization versus binding can differ in a polyclonal context.[1,47]

Using the deep mutational scanning system, we have mapped mutations that escape monoclonal antibodies targeting the RBD, NTD, and S2 domains. Although we only characterized a few antibodies here, this work is a first step toward generating similar data for much larger antibody panels. For the RBD, such data has proven valuable for interpreting the antigenic phenotypes of the new SARS-CoV-2 variants constantly emerging in the human population.[1,7] The ability to generate comparable data for other spike domains should enable better antigenic assessment of new variants, which can help inform vaccine design and strain selection.[48] In addition, our work on the S2 antibodies emphasizes the distinction between breadth on natural viruses and potential for escape within SARS-CoV-2: an antibody that cross-neutralizes MERS-CoV and all tested SARS-related viruses is still escaped by mutations already present in some SARS-CoV-2 variants. Therefore, direct prospective mapping of escape mutations[5,6] should be useful for informing the development of therapeutic antibodies targeting even regions of spike conserved across most natural viruses.

We also used the deep mutational scanning system to measure how mutations in spike affect its ability to mediate pseudovirus infection. These measurements complement existing deep mutational scanning datasets on how spike mutations affect the molecular phenotypes of ACE2 affinity,[13] membrane fusion,[11] and cell-surface expression.[10,11,13] None of these experimental measurements fully reflect how mutations affect actual viral fitness, which is an emergent property of many molecular phenotypes.[49,50] However, understanding how mutations affect molecular phenotypes is an important step toward interpreting evolutionary outcomes, and our pseudovirus-based system currently provides the deep mutational scanning measurements that best reflect natural selection on the spike of human SARS-CoV-2.

The deep mutational scanning system we describe here can also be straightforwardly extended to any virus with an entry protein amenable to lentiviral pseudotyping. This set of viruses includes other coronaviruses, influenza viruses, filoviruses, arenaviruses, and henipaviruses—all of which have receptor-binding and fusion proteins for which lentiviral pseudotyping provides a safe way to study cellular infection and antibody neutralization without requiring direct work with the actual pathogenic virus.[51–55] Deep mutational scanning of the entry proteins of all these viruses could provide valuable information for antigenic surveillance and vaccine design since these proteins are the dominant target of neutralizing antibodies. However, we note that data generated by such a system could in principle be used to inform the introduction of gain-of-function mutations into actual potential pandemic viral pathogens. We therefore suggest that advances in the high-throughput characterization of mutations to viral proteins should be coupled with thoughtful limits on any downstream experiments with actual replicating viruses[56] to ensure that safely generated information is used to benefit public health without creating new risks.

Limitations of the study

The full spike lentiviral-based deep mutational scanning system described in this paper has both limitations and advantages compared with other systems to identify SARS-CoV-2 antibody-escape mutations. Yeast-display-based RBD deep mutational scanning[1,8,13] is probably more high-throughput (e.g., Cao et al.[1] have applied it to thousands of antibodies) but is limited to only RBD antibodies and measures binding rather than neutralization. The passaging of spike-pseudotyped VSV in the presence of antibodies can effectively identify both individual escape mutations and combinations of mutations, and better simulates natural evolution,[39,57] but does not provide comprehensive maps because the mutations that arise in any given experiment are stochastic. The passaging of live SARS-CoV-2 has similar benefits and limitations to the spike-pseudotyped VSV[58] but also involves biosafety concerns. The system we describe here is comprehensive and applies to the full spike, but will only identify escape mutations introduced into the original library design.

An important caveat is that our system measures the effects of spike mutations on viral entry in a pseudotyped system in cell culture. Although this system captures many of the actual constraints on spike, it does not capture all of them as evidenced by the fact that some of the mutations we identify that increase viral entry in our system do not appear to be favorably selected during the actual evolution of human SARS-CoV-2. Therefore, it is important to keep in mind that the measurements that are generated by our system come from a simplified experimental surrogate for the full role of spike during infection and transmission in humans.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials Availability
  - Data and code availability

mammalian display-based deep mutational scanning. Interactive plots that enable mouseovers and show correlations among experiments are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/all_natural_enrichment_vs_dms.html. Interactive plots with correlations of enrichments during natural evolution calculated only from BA.1 sequences are at https://dms-vep.github.io/SARS-CoV-2_Omicron_BA.1_spike_DMS_mAbs/21K_natural_enrichment_vs_dms.html and have a lower correlation with the deep mutational scanning probably because a smaller dataset leads to less accurate estimates of enrichment during natural evolution.